Ethics of "Taste, Ties and Time"

"Just because you can do it doesn't mean you should."

Apparently the researchers of the "Taste, Times and Time" study who downloaded from Facebook data pertaining to a certain college's graduating class of 2009 were guided more by the can rather than by the should. The idea was to develop a new versatile network dataset. Seems well-intentioned and quite practical as well -- the Facebook data is already collected, sorted and collated in a single database, it needs only to be accessed and utilized for research. Cost-efficient? The answer depends on which costs you are willing to take into account.

The Facebook data includes information on individual's background (e.g. high school, hometown), demographics (e.g. birthday, gender), "interests," political views, and group affiliations, as well as on their cultural tastes (e.g. "favorite" books, movies, and music), pictures and network of friendships.[1] These were shared by individuals with Facebook without informed consent regarding the fact that the data can be used by researchers for secondary purposes. The users did not have an option to correct or delete any piece of information in the database.[2]

In addition, it was combined with the student housing records obtained from the college authorities, which again the students had no idea of being used for any purpose other than administrative purpose of the college. Does the cost of disrespecting an individual's autonomy feature in the cost analysis?

If yes, then not only did the researchers incur greater cost by not taking permission to download the profiles from the students, they actually went a step further and let undergraduate and graduate students research associates (RA) of the same college download the data. The researchers conceded that one RA might have different access to a student's profile than a different RA and being ''public'' or ''private'' on Facebook is merely relative to that particular RAs level of access.[3] The researchers overlooked that privacy is contextual.

One of the ways that people deal with the challenges to their privacy online is to employ multiple strategies for managing identity and reputation across different networks and transactions. Users bounce back and forth between different levels of disclosure depending on the context. People choose to engage using wide range of identifiers-screen name, actual name, and anonymity. So having agency over how you share what you share is crucial to privacy. [4] We often agree to go to Terms of services without fully understanding how the data would be handled, who would be the owner of the data or how it would be used.

Kaufman had further argued that the study did not harm anyone, ''what might hackers want to do with this information, assuming they could crack the data and 'see' these people's Facebook info?'', he argued. Kaufman's mention of ''hackers'', ''attacking'' the dataset,

1. Kevin Lewis et al., "Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com,"Social Networks 30, no. 4 (2008)
2. Michael Zimmer, ""But the Data Is Already Public": On the Ethics of Research in Facebook," Ethics and Information Technology 12, no. 4 (2010): , doi:10.1007/s10676-010-9227-5.
3. Zimmer, "But the Data Is Already Public"
4. "Danah Boyd: Privacy Challenges for Big Data (DataEDGE Conference 2012),"

and focusing on what someone might ''do'' with this information exposes a harm-based theory of privacy protection. It ignores the broader dignity-based theory of privacy. [5] Studies have shown that people's movements and activities change when they know they are or can be watched all the time. The threat of surveillance, real or imaginary invades personal space and hinders freedom of expression and creativity. [6]

What made these actions even more questionable is the research team's decision to publicly release part of its archive. While the team according to its research methodology took efforts to remove and/or personal identifiers, using only the publically available codebook for the dataset the college was correctly identified as the Harvard College by a privacy researcher. [7] Looking at the data a bit closer, for some nationalities represented by only one person, the identity of the involved individual becomes obvious. If data about favourites are further looked at, it will reveal information tied to identification. That such a scenario is indeed possible was evidenced by two researchers Arvind Narayanan and Vitaly Shmatikov. They studied the insufficiently anonymous information released by Netflix about nearly half-a-million customers as part of its $1 million contest to improve its recommendation system. Just weeks after the contest began the researchers identified several NetFlix users by comparing their "anonymous" reviews in the Netflix data to ones posted on the Internet Movie Database website. Revelations included identifying their political leanings and sexual orientation. [8] In fact an in-the-closet lesbian mother sued Netflix for privacy invasion, alleging the movie rental company made it possible for her to be outed as a consequence of the context. [9]

Revealing such sensitive information has wide range of consequences, embarrassment and insecurity being a few of those. Apple boss, Tim Cook proclaims that "history has shown us that sacrificing our right to privacy can have dire consequences. We still live in a world where all people are not treated equally. Too many people do not feel free to practice their religion or express their opinion or love who they choose." [10]

Instances like these show that collecting data indiscriminately and releasing them without sufficient anonymization carries with it the potential risk of harming the individuals who form points in the dataset. The basic principle of beneficence wherein risk minimization is of utmost importance is thus being violated, without any social good in sight.

Moreover the web by virtue of its digital memory and techniques of deep search will retain some ephemeral thought or idea one had randomly shared. While in a non-networked world it would have been forgotten with time or confined to a small set of people, it now has the potential of having being dug up years later with some major repercussions considering new acquaintances or employers. So the digital memory gets intricately twisted with the questions of privacy.

It can make sense for businesses to carry these activities to generate greater profit by delivering targeted advertising, there remains no argument for the social science researchers

5. Zimmer, "But the Data Is Already Public"
6. Danah Boyd: Privacy Challenges for Big Data (DataEDGE Conference 2012),"
7. Amabile, T., Goldfarb, P. and Brackfleld, S. (1990). Social influences on creativity: Evaluation, coaction, and surveillance. Creativity Research Journal, 3(1), pp.6-21.
8. How To Break Anonymity of the Netflix Prize Dataset
9. "Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims," Wired.com
10. Cook, T. (2015). Tim Cook: 'Sacrificing our right to privacy can have dire consequences'

to go this way especially in absence of substantial public benefit. Scientists obviously are held at a much greater regard than businessmen.

The scientists, from a legal perspective were not at fault when they conducted the study. They obtained permission from Facebook and the Harvard IRB. This is where ethics comes into play. Ethics is knowing the difference between what you have a right to do and what is right to do. While no-one involved technically broke the law, there definitely was lapse of judgement on the part of IRB when they failed to judge how the privacy of the people in the dataset could be compromised. While the researchers can feign ignorance, the people on the IRB panel who are experts on research ethics cannot absolve themselves. [11] The researchers also upheld transparency based accountability when they made their research along with the methodologies available in public domain but faltered when it came to law vs ethics.

While there has been no direct pitfall of this study, maybe in part because the group after criticism refrained from making the data public, it should be remembered that in an elusive search for greater good, this small specific set had to bear the burden.

Presented with such a dataset, I too would be faced with the question: can the means justify the end or can the ends justify the means? I would prefer not using this data since I know how it has been collected and that I cannot guarantee complete anonymization of it. I would be uncomfortable knowing my information can be distributed and utilized without my prior consent and I would assume that this viewpoint extends to people whose data is being used without their consent. The mere availability of data and participants' willingness to share it does not grant me an automatic right to use it freely.

---

11. Zimmer, "But the Data Is Already Public"

Bibliography

1. Lewis, Kevin, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. "Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com." *Social Networks* 30, no. 4 (2008): 330-42. doi:10.1016/j.socnet.2008.07.002.
2. Zimmer, Michael. ""But the Data Is Already Public": On the Ethics of Research in Facebook." *Ethics and Information Technology* 12, no. 4 (2010): 313-25. doi:10.1007/s10676-010-9227-5.
3. BerkeleyISchool. "Danah Boyd: Privacy Challenges for Big Data (DataEDGE Conference 2012)." YouTube. 2012. Accessed October 10, 2016. https://www.youtube.com/watch?v=aOIUYWW4uv8.
4. Amabile, T., Goldfarb, P. and Brackfleld, S. (1990). Social influences on creativity: Evaluation, coaction, and surveillance. *Creativity Research Journal*, 3(1), pp.6-21.Accessed October 10, 2016.
5. "ArXiv.org Cs ArXiv:cs/0610105." [cs/0610105] How To Break Anonymity of the Netflix Prize Dataset. Accessed October 10, 2016. http://arxiv.org/abs/cs/0610105.
6. "Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims." Wired.com. Accessed October 10, 2016. https://www.wired.com/2009/12/netflix-privacy-lawsuit/.

11. Zimmer, "But the Data Is Already Public"